



USPTO

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: The ACM Digital Library The Guide

THE ACM DIGITAL LIBRARY

[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used:

[set associative cache ways banks access time mux](#)

Found 39 of 215,737

Sort results by

 relevance
 Save results to a Binder

Display results

 expanded form
 Search Tips

 Open results in a new window

[Try an Advanced Search](#)
[Try this search in The ACM Guide](#)

Results 1 - 20 of 39

Result page: 1 [2](#) [next](#)

Relevance scale

1 [Track 3: high performance embedded architectures \(part 1\): Skewed caches from a](#)

[low-power perspective](#)

Mathias Spjuth, Martin Karlsson, Erik Hagersten

May 2005 **Proceedings of the 2nd conference on Computing frontiers CF '05**

Publisher: ACM Press

Full text available: [pdf\(214.02 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The common approach to reduce cache conflicts is to increase the associativity. From a dynamic power perspective this associativity comes at a high cost. In this paper we present miss ratio performance and a dynamic power comparison for set-associative caches, a skewed cache and also for a new organization proposed, the elbow cache. The elbow cache extends the skewed cache organization with a relocation strategy for conflicting blocks. We show that these skewed designs significantly reduce the co ...

Keywords: CAT, elbow, low-power, skewed caches

2 [A highly configurable cache for low energy embedded systems](#)

Chuanjun Zhang, Frank Vahid, Walid Najjar

May 2005 **ACM Transactions on Embedded Computing Systems (TECS)**, Volume 4 Issue 2

Publisher: ACM Press

Full text available: [pdf\(714.89 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Energy consumption is a major concern in many embedded computing systems. Several studies have shown that cache memories account for about 50% of the total energy consumed in these systems. The performance of a given cache architecture is determined, to a large degree, by the behavior of the application executing on the architecture. Desktop systems have to accommodate a very wide range of applications and therefore the cache architecture is usually set by the manufacturer as a best compr ...

Keywords: Cache, architecture tuning, configurable, embedded systems, low energy, low power, memory hierarchy, microprocessor

3 [A highly configurable cache architecture for embedded systems](#)

 Chuanjun Zhang, Frank Vahid, Walid Najjar
 May 2003 **ACM SIGARCH Computer Architecture News , Proceedings of the 30th annual international symposium on Computer architecture ISCA '03**, Volume 31 Issue 2

Publisher: ACM Press

Full text available:  pdf(302.87 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

Energy consumption is a major concern in many embedded computing systems. Several studies have shown that cache memories account for about 50% of the total energy consumed in these systems. The performance of a given cache architecture is largely determined by the behavior of the application using that cache. Desktop systems have to accommodate a very wide range of applications and therefore the manufacturer usually sets the cache architecture as a compromise given current applications, technolo ...

Keywords: architecture tuning, cache, configurable, embedded systems, low energy, low power, microprocessor

4 Power optimizations for cache memory: A way-halting cache for low-energy high-performance systems 

 Chuanjun Zhang, Frank Vahid, Jun Yang, Walid Najjar
 August 2004 **Proceedings of the 2004 international symposium on Low power electronics and design ISLPED '04**

Publisher: ACM Press

Full text available:  pdf(236.33 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Caches contribute to much of a microprocessor system's power and energy consumption. We have developed a new cache architecture, called a way-halting cache, that reduces energy while imposing no performance overhead. Our way-halting cache is a four-way set-associative cache that stores the four lowest-order bits of all ways' tags into a fully associative memory, which we call the halt tag array. The lookup in the halt tag array is done in parallel with, and is no slower than, the set-index decod ...

Keywords: cache design, low power techniques

5 Compiler-managed partitioned data caches for low power 

 Rajiv Ravindran, Michael Chu, Scott Mahlke
 June 2007 **ACM SIGPLAN Notices , Proceedings of the 2007 ACM SIGPLAN/SIGBED conference on Languages, compilers, and tools LCTES '07**, Volume 42 Issue 7

Publisher: ACM Press

Full text available:  pdf(432.35 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Set-associative caches are traditionally managed using hardware-based lookup and replacement schemes that have high energy overheads. Ideally, the caching strategy should be tailored to the application's memory needs, thus enabling optimal use of this on-chip storage to maximize performance while minimizing power consumption. However, doing this in hardware alone is difficult due to hardware complexity, high power dissipation, overheads of dynamic discovery of application characteristics, and ...

Keywords: embedded processor, hardware/software co-managed cache, instruction-driven cache management, low-power, partitioned cache

6 Interconnect design considerations for large NUCA caches 

Naveen Muralimanohar, Rajeev Balasubramonian
 June 2007 **ACM SIGARCH Computer Architecture News , Proceedings of the 34th**

 **annual international symposium on Computer architecture ISCA '07, Volume 35 Issue 2**

Publisher: ACM Press

Full text available:  pdf(325.80 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The ever increasing sizes of on-chip caches and the growing domination of wire delay necessitate significant changes to cache hierarchy design methodologies. Many recent proposals advocate splitting the cache into a large number of banks and employing a network-on-chip (NoC) to allow fast access to nearby banks (referred to as Non-Uniform Cache Architectures--NUCA). Most studies on NUCA organizations have assumed a generic NoC and focused on logical policies for cache block placement, movement ...

Keywords: cache models, interconnect, memory hierarchies, network-on-chip, non-uniform cache architecture

7 Cooperative Caching for Chip Multiprocessors

 Jichuan Chang, Gurindar S. Sohi

May 2006 **ACM SIGARCH Computer Architecture News , Proceedings of the 33rd annual international symposium on Computer Architecture ISCA '06, Volume 34 Issue 2**

Publisher: IEEE Computer Society, ACM Press

Full text available:  pdf(352.98 KB) Additional Information: [full citation](#), [abstract](#), [index terms](#)

This paper presents CMP Cooperative Caching, a unified framework to manage a CMP's aggregate on-chip cache resources. Cooperative caching combines the strengths of private and shared cache organizations by forming an aggregate "shared" cache through cooperation among private caches. Locally active data are attracted to the private caches by their accessing processors to reduce remote on-chip references, while globally active data are cooperatively identified and kept in the aggregate cache to re ...

8 The V-Way Cache: Demand Based Associativity via Global Replacement

 Moinuddin K. Qureshi, David Thompson, Yale N. Patt

May 2005 **ACM SIGARCH Computer Architecture News , Proceedings of the 32nd annual international symposium on Computer Architecture ISCA '05, Volume 33 Issue 2**

Publisher: IEEE Computer Society, ACM Press

Full text available:  pdf(231.93 KB) Additional Information: [full citation](#), [abstract](#), [cited by](#), [index terms](#)

As processor speeds increase and memory latency becomes more critical, intelligent design and management of secondary caches becomes increasingly important. The efficiency of current set-associative caches is reduced because programs exhibit a non-uniform distribution of memory accesses across different cache sets. We propose a technique to vary the associativity of a cache on a per-set basis in response to the demands of the program. By increasing the number of tag-store entries relative to the ...

9 The Vector-Thread Architecture

 Ronny Krashinsky, Christopher Batten, Mark Hampton, Steve Gerding, Brian Pharris, Jared Casper, Krste Asanovic

March 2004 **ACM SIGARCH Computer Architecture News , Proceedings of the 31st annual international symposium on Computer architecture ISCA '04, Volume 32 Issue 2**

Publisher: IEEE Computer Society, ACM Press

Full text available:  pdf(317.13 KB) Additional Information: [full citation](#), [abstract](#), [citations](#)

The vector-thread (VT) architectural paradigm unifies the vector and multithreaded compute models. The VT abstraction provides the programmer with a control processor and a vector of virtualprocessors (VPs). The control processor can use vector-fetch commandsto broadcast instructions to all the VPs or each VP can usethread-fetches to

direct its own control flow. A seamless intermixing of the vector and threaded control mechanisms allows a VT architecture to flexibly and compactly encode application ...

10 Memory hierarchy: Unified microprocessor core storage

Albert Meixner, Daniel J. Sorin

May 2007 **Proceedings of the 4th international conference on Computing frontiers CF '07**

Publisher: ACM Press

Full text available:  pdf(500.53 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The organization and management of microprocessor storage structures (e.g., L1 caches, TLBs, etc.) is critical to the performance and energy consumption of the microprocessor. We propose and develop the first microprocessor that can dynamically allocate storage to the structures that need it. First, we replace each existing structure with a dedicated micro-cache (μcache) that is smaller than is typical for that structure. With the smaller sizes, these structures can be made faster and le ...

Keywords: microarchitecture, power-efficiency, resource allocation, unified caching

11 Energy efficient architectures: Direct addressed caches for reduced power consumption

Emmett Witchel, Sam Larsen, C. Scott Ananian, Krste Asanović

December 2001 **Proceedings of the 34th annual ACM/IEEE international symposium on Microarchitecture MICRO 34**

Publisher: IEEE Computer Society

Full text available:  pdf(1.09 MB)  Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)
Publisher Site

A *direct addressed cache* is a hardware-software design for an energy-efficient microprocessor data cache. Direct addressing allows software to access cache data without a hardware cache tag check. These tag-unchecked loads and stores save the energy of a tag check when the compiler can guarantee an access will be to the same line as an earlier access. We have added support for tag-unchecked loads and stores to C and Java compilers. For Mediabench C programs, the compiler eliminates 16-76% ...

12 Addressing mode driven low power data caches for embedded processors

Ramesh V Peri, John Fernando, Ravi Kolagotla

June 2004 **Proceedings of the 3rd workshop on Memory performance issues: in conjunction with the 31st international symposium on computer architecture WMPI '04**

Publisher: ACM Press

Full text available:  pdf(337.52 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The size and speed of first-level caches and SRAMs of embedded processors continue to increase in response to demands for higher performance. In power-sensitive devices like PDAs and cellular handsets, decreasing power consumption while increasing performance is desirable. Contemporary caches typically exploit locality in memory access patterns but do not exploit locality information encoded in addressing modes used to access memory. We present two schemes that use locality information inherent ...

13 Register file and memory system design: Reducing register ports for higher speed and lower energy

I Park, Michael D. Powell, T. N. Vijaykumar

November 2002 **Proceedings of the 35th annual ACM/IEEE international symposium on Microarchitecture MICRO 35**

Publisher: IEEE Computer Society Press

Full text available: [pdf\(1.28 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)
[Publisher Site](#)

The key issues for register file design in high-performance processors are access time and energy. While previous work has focused on reducing the number of registers, we propose to reduce the number of register ports through two proposals, one for reads and the other for writes. For reads, we propose bypass hint to reduce register port requirements by avoiding unnecessary register file reads for cases where values are bypassed. Current processors are unable to avoid these unnecessary reads due ...

14 Survey of commercial parallel machines

Gowri Ramanathan, Joel Oren
 June 1993 **ACM SIGARCH Computer Architecture News**, Volume 21 Issue 3

Publisher: ACM Press

Full text available: [pdf\(1.64 MB\)](#) Additional Information: [full citation](#), [abstract](#), [citations](#), [index terms](#)

We have presented in this paper the survey of the parallel machines that are marketed today. The survey includes the latest machines available from Kendall Square Research, Thinking Machines Corporation, MasPar Computer Corporation, NCUBE Corporation, Sequent Computer Systems and Parsytec. We have provided the topology, architecture, cache coherence, synchronization and performance in MFLOPs for each of the machines subject to the availability of information.

15 Register Packing: Exploiting Narrow-Width Operands for Reducing Register File Pressure

Oguz Ergin, Deniz Balkan, Kanad Ghose, Dmitry Ponomarev
 December 2004 **Proceedings of the 37th annual IEEE/ACM International Symposium on Microarchitecture MICRO 37**

Publisher: IEEE Computer Society

Full text available: [pdf\(224.06 KB\)](#) Additional Information: [full citation](#), [abstract](#)

A large percentage of computed results have fewer significant bits compared to the full width of a register. We exploit this fact to pack multiple results into a single physical register to reduce the pressure on the register file in a superscalar processor. Two schemes for dynamically packing multiple "narrow-width" results into partitions within a single register are evaluated. The first scheme is conservative and allocates a full-width register for a computed result. If the computed result tu ...

16 Architecture -- memory hierarchy: Increasing cache capacity through word filtering

Prateek Pujara, Aneesh Aggarwal
 June 2007 **Proceedings of the 21st annual international conference on Supercomputing ICS '07**

Publisher: ACM Press

Full text available: [pdf\(786.66 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

With the increasing performance gap between processor and memory, it is essential that caches are utilized efficiently. However, caches are very inefficiently utilized because not all the excess data fetched into the cache, to exploit spatial locality, is accessed. Studies have shown that a prediction accuracy of about 95% can be achieved when predicting the to-be-referenced words in a cache block. In this paper, we use this prediction mechanism to fetch only the to-be-referenced data into th ...

Keywords: cache capacity, cache compression, cache miss rate, cache noise, cache organization

17 Reducing cache energy through dual voltage supply

 Vasily G. Moshnyaga

January 2001 **Proceedings of the 2001 conference on Asia South Pacific design automation ASP-DAC '01**

Publisher: ACM Press

Full text available:  pdf(170.15 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Due to a large capacitance and enormous access rate, caches dissipate about a third of the total energy consumed by today's processors. In this paper we present a new architectural technique to reduce energy consumption in caches. Unlike previous approaches, which have focused on lowering cache capacitance and the number of accesses, our method exploits a new freedom in cache design, namely the voltage per access. Since in modern caches, the loading capacitance operated on cache-hit is much I ...

18 Spatial computation

 Mihai Budiu, Girish Venkataramani, Tiberiu Chelcea, Seth Copen Goldstein

October 2004 **ACM SIGARCH Computer Architecture News , ACM SIGPLAN Notices , ACM SIGOPS Operating Systems Review , Proceedings of the 11th international conference on Architectural support for programming languages and operating systems ASPLOS-XI**, Volume 32 , 39 , 38 Issue 5 , 11 , 5

Publisher: ACM Press

Full text available:  pdf(573.00 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

This paper describes a computer architecture, *Spatial Computation* (SC), which is based on the translation of high-level language programs directly into hardware structures. SC program implementations are completely distributed, with no centralized control. SC circuits are optimized for wires at the expense of computation units. In this paper we investigate a particular implementation of SC: ASH (Application-Specific Hardware). Under the assumption that computation is cheaper than co ...

Keywords: application-specific hardware, dataflow machine, low-power, spatial computation

19 Scalable Load and Store Processing in Latency Tolerant Processors

 Amit Gandhi, Haitham Akkary, Ravi Rajwar, Srikanth T. Srinivasan, Konrad Lai

May 2005 **ACM SIGARCH Computer Architecture News , Proceedings of the 32nd annual international symposium on Computer Architecture ISCA '05**, Volume 33 Issue 2

Publisher: IEEE Computer Society, ACM Press

Full text available:  pdf(187.74 KB) Additional Information: [full citation](#), [abstract](#), [cited by](#), [index terms](#)

Memory latency tolerant architectures support thousands of in-flight instructions without scaling cycle-critical processor resources, and thousands of useful instructions can complete in parallel with a miss to memory. These architectures however require large queues to track all loads and stores executed while a miss is pending. Hierarchical designs alleviate cycle time impact of these structures but the CAM and search functions required to enforce memory ordering and provide data forwarding pl ...

20 On pipelining dynamic instruction scheduling logic

 Jared Stark, Mary D. Brown, Yale N. Patt

December 2000 **Proceedings of the 33rd annual ACM/IEEE international symposium on Microarchitecture MICRO 33**

Publisher: ACM Press

Full text available:  pdf(128.82 KB)

 ps(543.84 KB)Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#) Publisher Site

Results 1 - 20 of 39

Result page: [1](#) [2](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2007 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:

 [Adobe Acrobat](#) [QuickTime](#) [Windows Media Player](#) [Real Player](#)